University of Groningen

Mixture model analysis of complex samples

Wedel, M.; Hofstede, F. ter; Steenkamp, J.-B.E.M.

Mixture Model Analysis of Complex Samples

Michel Wedel[1]
University of Groningen

Frenkel ter Hofstede
Wageningen University and University of Groningen

Jan-Benedict E.M. Steenkamp
Catholic University of Leuven and Wageningen University

SOM theme B: Marketing and Networks

Januari 1997

Abstract
We investigate the effects of a complex sampling design on the identification of underlying classes from the sample using mixture models. A pseudo-likelihood approach is proposed and applied to obtain consistent estimates of class-specific parameters in the population. The effects of ignoring complex sampling designs are demonstrated empirically in the context of an international value segmentation study.

Keywords: Mixture models, Complex Samples, Stratified Samples, Pseudo-likelihood

---

We are grateful to Prof. W.A. Kamakura for providing comments on earlier drafts and ideas for this paper.

1. Introduction

In the last five to ten years, the mixture model approach has seen an impressive upsurge in interest in the classification, psychometrics and marketing literature. Mixture model applications have typically assumed that the subjects in the sample are drawn from the population using a simple random sampling procedure. However, in practice such random samples are not necessarily desirable and seem to be the exception rather than the rule. Data often serve more than one purpose, and not all the purposes of the data need to be fully specified at the time of collection. Since large scale data collection can be very expensive, various purposes are sought and the data are to be used for some time. Thus, it is unlikely that the sampling design chosen is optimal for all future purposes of the data.

The framework for sampling theory has been developed by Neyman (1934). He established the role of randomisation as the basis for sampling strategies, and introduced the ideas of stratification and the use of unequal selection probabilities. From this followed developments on multi-stage sampling, and a general theory of sampling. In probability samples, the selection probabilities of all elements in the population are known. Contrary to non-probability samples, probability samples allow for the projection of the sample estimates to the population, and enable the calculation of the precision of these estimates. Apart from simple random sampling, the most important probability sampling strategies are stratified sampling, cluster sampling, and two-stage sampling. We refer to such sampling procedures as <u>complex</u> sampling procedures. Good surveys use the structure of the population and employ sampling designs that incorporate stratification and clustering of the observations to yield more precise estimates. The theory of probability-weighted estimation for descriptive purposes (for example estimating population totals and means) is well established (e.g. Cochran 1977). On the other hand, probability-weighted estimation for analytic, model based purposes has received attention only fairly recently. The book by Skinner, Holt and Smith (1993) provides an overview of developments in this area. The emphasis in sample surveys has traditionally been on description. However, surveys are increasingly used for analytic purposes, including classification. Whereas in traditional inference for descriptive purposes the complexities in the sample design are often intimately connected to the specifics of the estimation procedures employed, the application of statistical methods for data-analysis often do not take the complexity of the sampling strategy into account.

In this paper we deal with the problem of how to identify unobserved classes from samples that arise from complex probability sampling strategies. We are concerned with statistical inferences about the underlying class-structure of the population, on the basis of data that are obtained using a complex sample design. This problem has to our knowledge not previously been dealt with in the literature. It arises from the fact that the conventional procedures for estimating class-level parameters using mixture models are based upon the assumption of simple random sampling and independent and identically distributed observations. We show that, if the data come from a complex probability sample, inferences on classes in the population can be made by applying pseudo maximum likelihood estimators. We empirically demonstrate the effects ignoring the sampling design in traditional ML based approaches.

2. Sample Design and the Mixture Approach

2.1 Pseudo-Maximum Likelihood Approach

Mixture models are traditionally estimated under the assumption of simple random sampling. Not explicitly accounting for a sampling strategy other than simple random sampling results in inconsistent and biased estimates. The approach to deal with complex designs is based on the so-called pseudo-maximum likelihood (PML) estimation approach and requires the knowledge of the selection probabilities for each of the final units selected in the sample. The development below for the mixture approach is based on Skinner (1989). The PML approach has been applied to several statistical models, but as far as we know not to mixture models.

We introduce the following notation:

n	= 1,...,N indicate primary sampling units ;
m	= 1,...,M indicate secondary sampling units ;
$N^{(p)}$	= number of units in the population;
M_n	= cluster size of cluster n;
N	= sample size;
$\mathbf{y}_n$	= (Kx1) vector of sample observations on unit n;
$\mathbf{Y}_n$	= (Kx1) vector of population values for unit n.
g	= 1,...,G indicate strata;
$N_g^{(p)}$	= number of units in stratum g of the population;
N_g	= number of elements in stratum g in the sample;

Assume a general sampling strategy that may involve combinations of more specific sampling schemes, for example, stratified and two-stage sampling. Suppose that the sampling strategy is such that a unit n in the sample has a probability of being selected of P_n (here we do not distinguish between primary and secondary units yet). A simple random sample will set $P_n = P$, for all n, for example. Complex samples require different values of P_n reflecting different probabilities of sampling units of different types. The conventional estimation approach under simple random sampling is to fit parametric mixture models via maximum likelihood estimation.

We will formulate the problem in a fairly general mixture model context. The data on subject n consist of K measurements of some variable Y: $\mathbf{y}_n$.

Assume the existence of S unobserved classes, with unknown proportions $_s$. Given a particular class, the observations are assumed to be distributed with probability-density function $f_s(\boldsymbol{y}_n|\phi_s)$, where $f_s(\cdot\ )$ is known to be one of the exponential family, and the parameter vector ϕ_s characterising class s is unknown. The exponential family includes many distributions that have useful applications, such as the normal, binomial, multinomial, negative binomial, exponential, poisson and gamma distributions (cf. McCullagh and Nelder 1989). The common properties of these distributions enable them to be studied simultaneously, rather than as a collection of seemingly unrelated cases. Within the unobserved classes a variety of possible data-generating mechanisms may be assumed. First, the ϕ_s may involve a single constant (or K constants), giving rise to standard mixtures of exponential family distributions (Titterington, Smith and Makov 1985). Second, if covariates $\mathbf{X}_n$ are measured for each n, a mixture of generalized linear models may be assumed to underlie the data, and the ϕ_s pertain to class-specific regression parameters (Wedel and DeSarbo 1995). Third, ϕ_s may involve stimulus locations and subject preference parameters in the case of mixtures of exponential family unfolding models (Wedel and DeSarbo 1996). In addition ϕ_s may include (known or unknown) nuisance parameters in the case of certain distributions in the exponential family (normal, negative binomial, etc.). Note that any of the parameters ϕ_s may be restricted to have the same value across classes. The unconditional distribution of the observations is formulated as:

$$f(\boldsymbol{y}_n\,|\,\phi) \ = \ \sum_{s=1}^{S} \ {}_s f_s(\boldsymbol{y}_n\,|\,\phi_s) \qquad (1)$$

The ML estimator of ϕ =($_s$, ϕ_s) maximizes the log-likelihood. The standard formulation of the log-likelihood applies under simple random sampling, in which each unit receives the same weight. The ML estimator solves the likelihood equations:

$$\sum_{n=1}^{N} \boldsymbol{J_n}(\phi) = \sum_{n=1}^{N} \frac{\partial \log f(\boldsymbol{y_n} \mid \phi)}{\partial \phi} = \mathbf{0} \tag{2}$$

Assume that taking the specific complex sample design into account, the $\mathbf{y}_n$ follow a model in which they are independent with the same p.d.f. as given in (1). Now given the sample design, the estimator for ϕ is obtained from the expectation of the score vector $\mathbf{T}(\phi)$, in the population:

$$\boldsymbol{T}(\phi) = \boldsymbol{E}[\boldsymbol{J_n}(\phi)] = \mathbf{0} \tag{3}$$

Equation (3) is a population version of the likelihood equations (2). Often a full ML procedure is intractable, since the expression for the likelihood under the complex sampling strategy depends on assumptions about the (unknown) relationships between the $\mathbf{y}_n$ and the sample design variables. However, a simple approach is to construct a consistent estimator for $\mathbf{T}(\phi)$, defined as:

$$\hat{\boldsymbol{T}}(\phi) = \sum_{n=1}^{N} {}_n \boldsymbol{J_n}(\phi). \tag{4}$$

The weights $_n$ are inverse proportional to the selection probabilities P_n and defined as $_n = \dfrac{N}{P_n \sum_{n=1}^{N} \frac{1}{P_n}}$, so that they sum to N across the sample. Solving equation (4) yields the so-called pseudo maximum estimator (PML) for ϕ. Complex sampling designs for which all selection probabilities are equal are “self weighting”, and the ML and PML estimators coincide. This is also the case if the distribution of $\mathbf{y}_n$ in (1) is completely independent of the sample design. This is, however, unlikely to occur in practice. Therefore, neglecting the sampling design for samples that are not selfweighting will lead to biased estimates.

It is assumed above that the mixture model is true in the population, which is assumed to be of infinite size. For many classification studies however, the mixture model is only a convenient approximation to the heterogeneity that exists in the “real world” and the parameters of the model are used to understand the approximate heterogeneity in a finite population. For such a finite population, we can define the population parameter ϕ as the solution of the likelihood equations over all units in the population. Thus, the mixture model is used as a working model to define the target parameters ϕ in the population. In this case we are only concerned with the distribution of ϕ due to the sampling design used. Under these much weaker conditions in which the mixture model is not assumed to be correct, the above PML estimation procedure remains valid (cf. Skinner 1989).

Below, we provide a few examples of the form of the selection probabilities in complex samples.

2.2 Stratified Samples

Stratified sampling is probably the most widely used complex sample design. It is assumed that the population is grouped into G strata. Each stratum, g, may arise from a combination of several stratification variables. Within stratum g, N_g subjects are sampled from the population. A mixture model is applied to the N (Kx1) observation vectors $\mathbf{y}_n$. If the distribution of $\mathbf{y}_n$ depends on the stratification variables, the ML estimates of the class-specific parameter estimates are not unbiased estimates of the population parameters. Likewise, the class sizes estimated from the sample with ML are not unbiased estimates of the sizes of the classes in the population. This is caused by the subjects composing the classes having unequal probabilities of being selected into the sample. The appropriate PML estimates of the parameters are weighted estimates obtained from equation (4), where the selection

probabilities equal: $P_n = \dfrac{N_g}{N_g^{(p)}}$, where g is the stratum from which respondent n comes. If the ratio of sample size and population size in each stratum is constant, $P_n = P$, so that the sample is selfweighting and the ML and PML estimators coincide.

2.3 Cluster Samples

If the units in the population occur naturally in clusters or primary units, cluster samples are often employed for reasons of cost reduction. Each primary sampling unit n comprises secondary units, indicated by $m=1,...,M_n$. A sample of primary units (n=1,...,N) is drawn, and observations on all secondary units in each primary unit are obtained ($M_n^{(p)} = M_n$) . Assume that classes need to be identified at the level of the secondary units, denoted by m (if classes are to be identified at the level of the primary units, the mixture model is applied to the (KM_n x1) vectors of observations on the primary units taking the selection probabilities for the primary units into account). For example, when the primary units all have the same size M and when the sample drawn from them is a simple random sample, then the weights of the secondary units are equal to

$P_{nm} = \frac{1}{NM}$ and the PML and ML estimators coincide. Note that in this situation in equation (4) the summation over n is replaced by a summation over n and m. If the primary units have unequal sizes, M_n, and they are drawn with random sampling the selection probabilities for the secondary units equal

$P_{nm} = \frac{M_n}{\sum_{n=1}^{N^{(p)}} M_n}$, where n is the primary unit from which m comes. If the primary units are of unequal size and drawn with probabilities proportional to their size, the sample is self-weighting. Cochran (1977) describes alternative procedures for obtaining cluster samples from which the selection probabilities can be easily derived.

2.4 Two-Stage Samples

In two-stage sampling methods, a sample of size N is drawn from all primary units in the population, and from each primary unit a sample of secondary units of size $M_n \leq M_n^{(p)}$, is drawn. Two-stage sampling procedures are often cheaper than and have higher statistical efficiency than cluster samples, while the latter may be infeasible when the primary units are large. The results for the selection probabilities for various selection strategies for the primary and secondary units are derived from the standard results provided in Cochran (1977). For example, if the primary units in the population have the same size ($M_n^{(p)} = M^{(p)}$), the N primary units and the M secondary units in the sample are selected by simple random sampling, and a constant fraction $M/M^{(p)}$ is sampled from each primary unit, the sample is self-weighting and all selection probabilities equal $P_{nm} = \frac{1}{NM}$. More typical are situations in which the primary units vary in size. Then the secondary units may be selected either with equal probabilities, or with probabilities proportional to size. Various available strategies involve different sampling and sub-sampling methods. If the secondary units and primary units are selected by simple random sampling, the sizes of the primary units differ, and a constant number M is sampled from each primary unit the selection probabilities are $P_{nm} = \frac{M_n^{(p)}}{\sum_{n=1}^{N^{(p)}} M_n^{(p)}}$.

In the situation that the sampling fraction within each primary unit is constant: $M_n/M_n^{(p)}=f_2$, say, the sampling strategy is self-weighting. Further results can be derived from Cochran (1977).

3. Statistical Inference

3.1. Asymptotic Standard Errors of the Estimates

Under typical regularity conditions the ML estimators are asymptotically normal. A consistent estimator of the asymptotic covariance matrix of the estimates is the inverse of the observed Fisher information matrix (e.g. Titterington, Smith and Makov 1985). The pseudo-log likelihood estimator in expression (4), however, is not efficient (it does not achieve the minimum variance among all possible estimators). The reason for this is that the optimal weighting of the units is the weighting obtained from the standard maximum likelihood function, and introducing the selection probabilities as weights decreases the efficiency of the estimator. (This points to the advantages of using self-weighting samples for the purpose of mixture model estimation, since the estimates have minimum variance because the weights cancel.) A robust estimator of the asymptotic variance is in this situation provided by White (1982), and Royall (1986), which in the general case of a stratified multi-stage sample is:

$$_{\boldsymbol{PML}}(\hat{\phi}) = \boldsymbol{H}(\hat{\phi})^{-1}\boldsymbol{V}(\hat{\phi})\ \boldsymbol{H}(\hat{\phi})^{-1}, \tag{5}$$

where

$$\boldsymbol{H}(\phi) = \frac{\partial \hat{\boldsymbol{T}}(\phi)}{\partial \phi} \tag{6}$$

is the matrix of second order partial derivatives, and

$$\boldsymbol{V}(\hat{\phi}) = \sum_{g=1}^{G} \frac{N_g}{N_g - 1} \sum_{n=1}^{N_g} (\boldsymbol{T}_{\boldsymbol{gn}}(\hat{\phi}) - \bar{\boldsymbol{T}}_{\boldsymbol{g}}(\hat{\phi}))(\boldsymbol{T}_{\boldsymbol{gn}}(\hat{\phi}) - \bar{\boldsymbol{T}}_{\boldsymbol{g}}(\hat{\phi}))' \tag{7}$$

with :

$$\boldsymbol{T}_{\boldsymbol{gn}}(\hat{\phi}) = \sum_{m=1}^{M_{gn}} {}_{n}\frac{\partial \ln f(\boldsymbol{y}_{\boldsymbol{nmg}} \mid \phi)}{\partial \phi} \tag{8}$$

where M_{gn} denotes the number of units in cluster n within stratum g.

3.2 Criteria for Selecting the Number of Classes

When applying mixture models the true number of classes, S, is mostly unknown and has to be inferred from the data. The problem of identifying the number of classes in mixture models has as yet not seen an entirely satisfactory statistical solution. Suppose one wishes to test the null-hypothesis (H_0) of S segments against the alternative hypothesis (H_1) of S+1 segments. The standard likelihood ratio test statistic is not applicable, because it is not asymptotically distributed as ^{2}. In testing for the number of components in a mixture model this asymptotic distribution is not valid, since H_0 corresponds to a boundary of the parameter space for H_1, a situation that violates the required regularity conditions (cf. Aitkin and Rubin 1985). Recently, Böhning, Dietz, Schaub, Schlattmann and Lindsay (1994) investigated the distribution of the LR test of S=1 versus S=2 mixtures of exponential families. They found that its limiting distribution is not well approximated by the conventional ^{2} distribution, and that the deviation is to be distribution specific.

Information criteria are therefore frequently used for investigating the number of classes. These criteria impose a penalty upon the log-likelihood which is related to the number of parameters estimated:

$$C(S) = -2\ln L(\hat{\phi}|S) + Qd \qquad (9)$$

Here, Q is the number of parameters estimated and d is some constant. That number of segments is selected, where the statistics reach a minimum value. The classical Akaike's Information Criterion, AIC, arises when d=2. For the Bayesian Information Criterion, BIC, d=ln(N) and for the Consistent Akaike's Information Criterion, CAIC, d=ln(N+1). These two criteria impose an additional sample size penalty upon the log-likelihood. The information theoretic measure ICOMP, is based on the properties of the estimated information matrix $\hat{\boldsymbol{I}}(\phi)$**:** $d \;=\; -\ln\; tr[\hat{\boldsymbol{I}}(\phi)^{-1}] - \frac{1}{Q}\ln\; \det[\hat{\boldsymbol{I}}(\phi)^{-1}]$, (10)

ICOMP penalizes the likelihood more when more parameters are estimated, but also when the model becomes less well identified due to an increasing number of parameters, in which case the term involving the determinant of the information matrix increases.

A problem with these criteria is that they depend on the likelihood and therefore rely on the same properties as the likelihood ratio test. Therefore they can be used only as indicative for the number of segments. In addition a complex sample design that is not taken into account in the formulation of the likelihood will affect the determination of the number of classes. We therefore propose that the information and entropy statistics should be based on the pseudo-log-likelihood:

$$\ln PL(\phi|S) \;=\; \sum_{n=1}^{N}\sum_{m=1}^{M_n} {}_{mn}\ln\sum_{s=1}^{S} {}_{s} f_s(\boldsymbol{y}_{\boldsymbol{mn}}|\phi_s) \tag{11}$$

Thus for complex samples lnPL(ϕ|S) replaces lnL(ϕ|S) in the equations (9) and (12) above, and **I**(ϕ) in (10) is defined as in (6).

4. Illustration

4.1. Data

We illustrate the PML procedure in a stratified European value segmentation study. The effects of taking the sampling design into account are demonstrated by comparing the PML estimates to the ML estimates. The purpose of the study is to investigate the existence of pan-European value segments, i.e. segments that transcend national borders. A “value” is defined in consumer psychology as an enduring belief that a specific state of existence or mode of conduct is preferred for living one’s life (cf. Rokeach 1973). The data were collected in 1996 in six West-European countries. The European sample was stratified by country. From each country a sample of approximately the same size was drawn, although the countries differ substantially in population size. This is a standard procedure in international value research. Specifically, the countries were (sample sizes/ weights in parentheses): Belgium (648/ 0.231), Germany (673/ 2.020), Great-Britain (623/ 1.356), France (694/ 1.259), the Netherlands (646/ 0.380), and Spain (616/ 0.600).

Kahle’s “List of Values” (LOV) method was employed to assess value-systems of respondents (Kahle 1986). The LOV typology is related to social distinction theory. It distinguishes between external and internal values, and deals with the importance of interpersonal relations, and personal and a-personal factors in value fulfilment. The LOV instrument is composed of nine values. These were ranked in a paper and pencil task by the respondents (back-translation methods were used in order to ensure a similar content of the statements in the languages involved, Brislin 1970). The nine values are:

LOV-1. Fun and enjoyment in life;
LOV-2. Warm relationships with others;
LOV-3. Self-fulfilment;
LOV-4. Being well respected;
LOV-5. Sense of belonging;
LOV-6. Excitement;
LOV-7. A sense of accomplishment;
LOV-8. Security;
LOV-9. Self-respect;

Thus, the data $\mathbf{y}_n$ consist of the order of the above values, for each of the 3900 respondents, where the lower the rank number, the more important a value is for a person.

4.2. The Model

To identify latent value segments a mixture of rank order multinomial logit models is used, proposed by Kamakura and Mazzon (1991). The model is a mixture model extension of Thurstone's (1927) law of comparative judgement and assumes that the observed value rankings are error-perturbed observations of the unobservable value utilities of each individual. The model is based on utility maximization theory and identifies segments on the basis of the entire value ranking provided by the subjects.

Assume the existence of S unobserved value segments. Individuals belonging to segment s share the same value system, represented by a set of unobserved utilities, U_{ks}, with k=1,...,K denoting the values. Let Y_{nt} denote the value label on rank order position t, t=1,...,T (T=K=9 for LOV). Given segment s, the probability of observing a value ranking $\{Y_{n1}, Y_{n2}, ..., Y_{nT}\}$ may be expressed in terms of the utilities as:

$$Prob(U_{Y_{n1}} \geq U_{Y_{n2}} \geq ... \geq U_{Y_{nT}}). \tag{12}$$

If the utilities are considered random with a standard Weibull distribution of the error component, this leads to the multinomial logit:

$$p_{k|s}(\boldsymbol{U_s}) = \frac{\exp[U_{k|s}]}{\sum_{t \in R_k} \exp[U_{t|s}]} \,, \tag{13}$$

with R_k the set of values ranked higher or equal to k. For reasons of identification, the utility of the last value (K=9) is set to zero. The pseudo-log-likelihood is:

$$L(\boldsymbol{y_n}\,|\,\boldsymbol{U}) = \sum_{n=1}^{N} {}_n \ln \sum_{s=1}^{S} {}_s \prod_{k=1}^{K} p_k(\boldsymbol{U_s})^{y_{nk}} \;, \tag{14}$$

with ${}_n = \dfrac{N_g^{(p)}\,N}{N_g\,N^{(p)}}$. The log-likelihood is obtained from equation (14) by setting P_n=1 for all n. The models are estimated using a weighed Quasi-Newton gradient search to minimize (16), using the Broyden, Fletcher, Goldfarb and Shanno procedure implemented in the GAUSS package (Aptech 1995).

4.3. Results

In order to illustrate the PML approach for mixture models, we estimate the model on these LOV data, using both the ML and the PML approach. Starting with S=1, S is increased until the AIC, CAIC, BIC and/or ICOMP show a minimum. In order to overcome problems of local optima, each model is estimated from 10 sets of random starting values. Table 1 shows the values of the log-likelihood, and the model selection criteria for ML estimation. Table 2 shows the statistics for PML estimation. With ML, BIC, CAIC and ICOMP all reach a minimum at S=7. Since AIC tends to overstate the number of classes, the seven class ML solution appears to be a reasonable representation of the data. For PML, BIC, CAIC and ICOMP and indicate that the S=5 solution is most appropriate.

The first important finding is that accounting for stratification using PML estimation may yield a different number of classes than when the stratified sample is ignored. Accounting for the stratified sampling procedure resulted in more parsimonious models (the S=5 solution has 18 parameters less than the S=7 solution). BIC, CAIC and ICOMP are consistent in their identification of the number of classes for ML and PML estimation, while AIC tends to indicate a too large number of classes.

[INSERT TABLES 1 AND 2 HERE]

Next, we inspect the ML and PML estimates. We focus on the S=5 solution, since that is the best approximation to the data, when the appropriate weighting is applied. The S=7 results are provided for completeness.

The prior probabilities (or class proportions) are presented in Tables 3 and 4. Table 3 shows that the S=5 ML and PML estimates are markedly different. First, the aggregate segment proportions ($_s$) are quite different between ML and PML. For PML, Segment 4 is much smaller (3%) and segment 5 much larger (38%) than the corresponding classes in the ML solution (12% and 28%). It appears that ML has a tendency to identify classes with more equal sizes. Given that PML does take the sample design into account, the comparison shows that the ML estimates are severely biased due to their negligence to take the sample design into account. Further, the PML and ML solutions show some substantial differences of the estimated class proportions per country. For example in the PML solution Germany has a particularly high proportion of Segment 2 (33%), while for ML this proportion is much lower (5%). On the contrary, for ML Segment 4 is substantial (43%) in Germany, while it is much smaller for PML (14%).As another example, PML-Segment 2 has a proportion of 2% in France, while ML-Segment 2 is 20% in that country.

For the S=7 solutions there are also some notable differences in the ML and PML estimates. Segments 1, 2, 3, 4 and 6 are larger for PML, the other segments are larger for ML. For the sizes of the segments per country, there are differences in particular for Segments 4 and 5 in Germany, for Segments 5 and 7 in Great Britain, for Segments 4 and 5 for France, and for Segments 5 and 7 for Spain, between ML and PML.

[INSERT TABLES 3 AND 4 HERE]

Tables 5 and 6 depict the value-utility estimates obtained with ML and PML for the S=5 and S=7 solutions, respectively. The S=5 estimates show some marked differences between PML and ML. For example for Segment 2, the ML procedure predicts a high value ranking for Self-fulfilment (LOV-3), while the PML procedure predicts a low ranking. On the other hand, PML predicts a high importance for Sense of Belonging (LOV-5) and Self-respect (LOV-8) in Segment 2, while low importances are predicted by the ML estimates. Other large differences between ML and PML occur in Segment 4 for Warm relationships with others (LOV-2), Self-fulfilment (LOV-3), Being well respected (LOV-4), Sense of belonging (LOV-5), and Excitement (LOV-6). In addition differences in the estimates are observed for Self-fulfilment (LOV-3) and for A sense of accomplishment (LOV-7) in Segment 3, and for Security (LOV-8), in Segment 1. In general, inspection of Table 5 reveals quite some differences. Table 6 shows that the PML and ML estimates for the S=7 solution show differences in particular for Segments 5 and 6. Overall, the differences are less striking than for the S=5 solution.

[INSERT TABLES 5 AND 6 HERE]

Finally, we show the extent to which the actual classification of the subjects in the sample corresponds between ML and PML. For that purpose, we "de-fuzzified" the classification obtained from the mixture models by assigning all subjects to that Segment for which the posterior probability of classification was largest (if ϕ were known this would be the optimal Bayes rule of classification, see McLachlan and Basford 1988, p. 11). Then, the memberships in the ML and PML segments were cross-tabulated. Tables 7 and 8 show these cross-tabulations for the S=5 and S=7 solutions respectively.

[INSERT TABLES 7 AND 8 HERE]

Table 7 shows that the classification of subjects into the five value-segments on the basis of the posteriors is quite different for ML and PML. The assignment of subjects to Segments 1 and 5 is quite similar, ML predicting close to 100% of the memberships correct, relative to PML. However, for segment 4 ML suffers from substantial misclassification relative to PML (23%). Moreover, none of the subjects in PML Segment 2 is assigned to the same segment with ML. This explains the very large differences in value utility estimates for Segment 2 reported above. A similar picture emerges for the S=7 solution in Table 8.

6. Conclusions

The purpose of this study is to investigate the effects of sample design on the standard maximum likelihood estimation of mixtures. To our knowledge, problems due to complex sample design have not previously been raised in the classification literature. The contribution of this study is to show how a relatively simple pseudo-maximum likelihood estimation procedure can be applied to finite mixtures of distributions in the exponential family, where several models may describe the expectations of the observations in each class, such as the standard mixture, the generalized linear and generalized nonlinear models. In addition we proposed to take the sample design into account in the likelihood- and entropy-based model selection criteria. In an empirical application, we demonstrated the effects of the PML approach for mixture models, and show that the estimates of the number of classes, the class proportions and the class level parameters may be severely biased when using ML instead of PML.

References

AITKIN, M. and RUBIN, D.B. (1985). Estimation and Hypothesis Testing in Finite Mixture Distributions, *Journal of the Royal Statistical Society, Series B*, 47, 67-75.

APTECH (1995). *Constrained maximum likelihood, GAUSS manual*. Aptech systems, Maple Valley.

BOHNING, D., DIETZ, E., SCHAUB, R., SCHLATTMANN, P., and LINDSAY, B.G. (1994). The Distribution of the Likelihood Ratio for Mixtures of Densities for the One Parameter Exponential Family. *Annals of the Institute of Statistics and Mathematics*, 46 (3), 373-388.

BRISLIN, R. (1970), Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology*, 1, 185-216.

COCHRAN, W.G. (1977). *Sampling Procedures.* New York :John Wiley.

KAHLE, L. (1986). The Nine Nations of North America and the Value Basis of Geographic Segmentation. *Journal of Marketing*, 50, 37-47.

KAMAKURA, W.A. and MAZZON, J.A. (1991), Value Segmentation: A Model for the Measurement of Values and Value Systems, *Journal of Consumer Research*, 18, 208-218.

MCCULLAGH, P. and NELDER, J.A. (1989). *Generalized Linear Models*. New York: Chapman and Hall.

MCLACHLAN, G.J. AND BASFORD K.E. (1988). *Mixture Models*. New York: Marcel Dekker.

NEYMAN J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, 97, 558-606.

ROKEACH, M. (1973). *The Nature of Human Values.* New York: Free Press.

ROYALL, R.M. (1986). Model Robust Confidence Intervals Using Maximum

Likelihood Estimators. *International Statistical Review*, 54, 221-226.
SKINNER, C.J., HOLT, D. AND SMITH, T.M.F (1989), *Analysis of Complex Surveys*. New York: John Wiley.
SKINNER, C.J. (1989), Domain Means, Regression and Multivariate Analysis. In: *Analysis of Complex Surveys,* Skinner, C.J., Holt, D. and Smith, T.M.F (eds.) New York: John Wiley, pp. 59-87.
TITTERINGTON, D.M., SMITH, A.F.M. AND MAKOV, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley.
THURSTONE, L.L. (1927), A Law of Comparative Judgment. *Psychological Review*, 34, 273-286.
WEDEL, M. AND DESARBO, W.S. (1995), A Mixture Likelihood Approach for Generalized Linear Models. *Journal of Classification*, 12 (1), 21-56.
WEDEL, M. AND DESARBO, W.S. (1996). An Exponential Family Mixture Unfolding Methodology. *Journal of Business and Economic Statistics*, 14 (4), 447-459.
WHITE, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50, 1-25.

Table 1

Model selection criteria for ML estimation[1] .

S	-ln-L	AIC	BIC	CAIC	ICOMP
1	45458.56	90933.13	90983.28	90991.54	90937.34
2	44892.24	89818.50	89925.06	89933.33	89842.63
3	44734.87	89521.75	89684.74	89693.01	89574.28
4	44530.97	89131.94	89351.35	89359.62	89213.08
5	44454.78	88997.56	89273.39	89281.66	89109.48
6	44397.70	88901.41	89233.65	89241.92	89056.67
7	44349.42	88822.95	**89211.51**	**89219.78**	**89034.50**
8	**44336.46**	**88814.92**	89260.00	89268.27	89322.11

[1] Boldface type indicates the minimum value across S=1 to S=8.

Table 2

Model selection criteria for PML estimation[1].

S	- ln-PL	AIC	BIC	CAIC	ICOMP
1	45458.56	90933.13	90983.28	90991.54	90937.34
2	44892.25	89818.49	89925.06	89933.33	89842.63
3	44734.88	89521.75	89684.74	89693.01	89576.66
4	44530.97	89131.94	89351.34	89359.61	89211.57
5	44471.89	89031.78	**89307.61**	**89315.88**	**89163.49**
6	**44452.25**	**89010.50**	89342.74	89351.01	89601.36

[1] Boldface type indicates the minimum value across S=1 to S=6.

Table 3

Prior probabilities for S=5 PML and ML solutions, aggregate and for each country.

Class	BE	DL	GB	FR	NL	SP	Total
PML 1	0.387	0.246	0.278	0.542	0.385	0.167	0.338
PML 2	0.038	0.325	0.115	0.020	0.047	0.022	0.095
PML 3	0.202	0.068	0.192	0.074	0.213	0.212	0.158
PML 4	0.004	0.138	0.015	0.002	0.004	0.009	0.029
PML 5	0.370	0.223	0.401	0.363	0.351	0.591	0.380
ML 1	0.326	0.257	0.231	0.435	0.350	0.123	0.290
ML 2	0.139	0.053	0.116	0.198	0.091	0.125	0.121
ML 3	0.236	0.102	0.240	0.090	0.246	0.251	0.191
ML 4	0.034	0.431	0.122	0.020	0.044	0.035	0.116
ML 5	0.266	0.157	0.291	0.258	0.269	0.467	0.282

Table 4

Prior probabilities for S=7 PML and ML solutions, aggregate and for each country.

Class	BE	DL	GB	FR	NL	SP	Total
PML 1	0.206	0.070	0.176	0.083	0.224	0.201	0.158
PML 2	0.128	0.055	0.120	0.173	0.083	0.144	0.117
PML 3	0.035	0.321	0.118	0.018	0.041	0.022	0.094
PML 4	0.322	0.235	0.223	0.442	0.338	0.119	0.283
PML 5	0.067	0.022	0.106	0.015	0.052	0.137	0.065
PML 6	0.005	0.149	0.018	0.004	0.006	0.011	0.033
PML 7	0.237	0.148	0.238	0.266	0.255	0.366	0.250
ML 1	0.194	0.059	0.179	0.065	0.200	0.207	0.148
ML 2	0.108	0.042	0.104	0.149	0.071	0.129	0.101
ML 3	0.040	0.349	0.128	0.021	0.049	0.026	0.103
ML 4	0.290	0.181	0.203	0.363	0.305	0.119	0.246
ML 5	0.087	0.092	0.051	0.138	0.083	0.030	0.081
ML 6	0.002	0.110	0.012	0.001	0.002	0.006	0.023
ML 7	0.279	0.167	0.322	0.263	0.289	0.482	0.298

Table 5

Value utilities in each of five segments for S=5 PML and ML solutions

Class	LOV-1	LOV-2	LOV-3	LOV-4	LOV-5	LOV-6	LOV-7	LOV-8
PML 1	2.324	1.140	-0.350	-0.119	-0.980	-1.228	-0.301	-0.163
PML 2	1.140	1.016	-1.569	-1.193	1.249	-2.276	-1.158	1.093
PML 3	-0.621	1.903	-0.849	0.464	0.071	-2.350	0.114	0.133
PML 4	0.874	-2.597	-2.158	-2.425	0.492	-3.270	-1.525	0.912
PML 5	-1.197	-0.664	-1.077	-0.769	-1.988	-2.928	-0.808	-0.356
ML 1	2.746	1.365	-0.532	-0.098	-0.868	-1.264	-0.377	0.009
ML 2	0.388	0.093	0.383	-0.567	-1.690	-1.723	-0.084	-0.876
ML 3	-0.579	1.808	-0.851	0.292	0.009	-2.350	-0.005	0.137
ML 4	0.786	-0.354	-1.394	-1.301	0.895	-2.231	-1.080	0.994
ML 5	-1.573	-0.929	-1.432	-0.866	-2.310	-3.345	-1.051	-0.265

Table 6

Value utilities in each of seven segments for S=7 PML and ML solutions

Class	LOV-1	LOV-2	LOV-3	LOV-4	LOV-5	LOV-6	LOV-7	LOV-8
PML	-0.307	2.077	-0.893	0.515	0.022	-2.184	0.083	0.166
PML	0.161	0.113	0.552	-0.552	-1.517	-1.812	0.050	-0.808
PML	0.903	0.860	-1.556	-1.306	1.227	-2.378	-1.220	1.009
PML	2.817	1.289	-0.451	-0.088	-0.900	-1.220	-0.348	-0.045
PML	-3.799	-0.368	-1.122	-0.540	-1.167	-5.321	-0.500	-0.388
PML	0.916	-2.336	-1.938	-2.139	0.446	-2.993	-1.317	0.862
PML	-1.208	-0.910	-1.523	-0.898	-2.470	-3.222	-1.175	-0.178
ML 1	-0.740	1.928	-0.876	0.481	0.062	-2.440	0.169	0.179
ML 2	0.114	0.181	0.691	-0.541	-1.529	-1.822	0.120	-0.849
ML 3	1.017	0.880	-1.471	-1.180	1.219	-2.234	-1.123	1.048
ML 4	2.489	1.701	-0.442	0.054	-0.774	-1.239	-0.327	-0.157
ML 5	4.129	-0.142	-0.694	-0.614	-1.551	-1.508	-0.450	0.347
ML 6	0.512	-3.104	-2.329	-2.782	0.471	-3.664	-1.602	0.961
ML 7	-1.542	-0.851	-1.455	-0.864	-2.235	-3.308	-1.107	-0.272

Table 7

Cross-classification of membership between S=5 PML and ML solutions

	ML 1	ML 2	ML 3	ML 4	ML 5	total
PML 1	1197	157	28	34	4	1420
PML 2	77		33	276		386
PML 3		9	529		1	539
PML 4				109		109
PML 5	11	188	122	48	1077	1446
total	1285	354	712	467	1082	3900
%	93.2	0.0	74.3	23.3	99.5	74.7

Table 8

Cross-classification of membership between S=5 PML and ML solutions

	ML 1	ML 2	ML 3	ML 4	ML 5	ML 6	ML 7	total
PML 1	445	2	13	62			55	577
PML 2		277		20	22		24	343
PML 3	1		365	2		1	5	374
PML 4		13	37	1042	153		4	1249
PML 5	85						220	305
PML 6		1	10		30	79	5	125
PML 7		2	1	12	61	1	850	927
total	531	295	426	1138	266	81	1163	3900
%	83.8	93.9	85.7	91.6	0.0	97.5	73.1	78.4